Rozsáhlé databáze osobních informací

Vašek Matyáš

PV080

Agregace dat

- Seskupování (osobních) dat do rozsáhlých databází. Agregace (z angl. *aggregation*).
- Tímto kombinováním dat o určité citlivosti lze získat informace daleko citlivější, které jinak spadají do kategorie s vyššími požadavky na ochranu.

Zákon o ochraně osobních údajů (101/2000 Sb.) – Povinnosti správce

Mj. zákon říká:

- nesdružovat osobní údaje, které byly získány k rozdílným účelům, pokud zvláštní zákon nestanoví jinak

Žadatel o investici

- Chodil roky ke stejnému obvodnímu lékaři.
- Uzavřel před měsícem vysokou živ. pojistku.
- V minulém čtvrtletí byl u specialisty.
- Přede dvěma měsíci změnil obvodního lékaře.

Odvození (Inference, i angl.)

- Odvození informací o vyšší citlivosti zpracováním a analýzou skupiny informací o nižší citlivosti.

nebo

- Nepřímý přístup k informacím bez přímého přístupu k datům, která tyto informace reprezentují.

Příklad politiky klinických IS, British Medical Association

- Musí být zavedena účinná opatření proti agregaci osobních zdravotních informací.
- Pacienti, k jejichž seznamu řízení přístupu má být přidána další osoba, musí být zvlášť upozorněni, pokud již tato osoba má přístup ke zdravotním informacím velkého množství lidí.

Co když máte informace o finanční situaci a zdrav. stavu

1. Přítele/kyně, resp. manžela/ky.
2. Spolupracovníka, nadřízeného...
3. Všech studentů/zaměstnanců FI.
4. Všech obyvatel místa, kde žijete.
5. Všech klientů určité firmy (banky, zdravotní pojišťovny...).
6. Všech (většiny) občanů.

Pravděpodobnost neoprávněného použití

- Počet osob, které mají k informacím přístup (operátoři, uživatelé systému ap.).
- Hodnota informací.
  - Výše trestu těm, kdo data jiných řádně neohlídali a spolupodíleli se tak na jejich úniku.
  - Výše trestu těm, kdo s nimi neoprávněně manipulují.
  - Úroveň ochranných mechanismů.

Řešení?

- U menších souborů osobních dat provádět agregaci jen v nutných případech.
- U větších souborů neprovádět agregaci.

- Statistické databáze!

Statistické databáze

- Obsahují citlivé údaje o jednotlivcích.
- Jejich využití má být **jen** pro statistické dotazy k vytvoření obrazu o celkových potřebách obyvatelstva a formulování (vládní) politiky.
  - podpora církví, regionů/měst atd.
- Výsledky dotazů v takovýchto databázích nesmějí poskytnout údaje o jednotlivcích.

Studium statistických databází

- USA, 70. léta, databáze ze sčítání lidu.
- Dorothy Denning
  - Studium používaných způsobů pro formulaci dotazů a získávání odpovědí.
  - Ty povolovaly (netriviální!) dotazy, které umožnily získat údajně tajné informace o jednotlivci.
  - ☺ Údajně nedůvěra ve zjistění Denningové – dokud nezjistila plat svého šéfa sérií legitimních dotazů.

Příklad kritického dotazu

Kolik je měst s 15-16 000 obyvatel

& s muži, evangelíky, slovenské nár., 36-40 let

& jejich ženy, 28-30 let žijí mimo toto město

& 2 děti do 10 let žijí s těmito ženami

& 1 dítě nad 18 žije s těmito muži

& muž žije ve vlastním domě, plocha nad 200m^2
a domácnost má/používá aspoň 2 automobily.

Kompromitace databáze

- Výsledkem série dotazů je jeden záznam
  - Databáze byla pozitivně kompromitována
- Následný pokus o získání dalších informací
  - Výsledkem je buď 1 nebo 0 záznamů
  - Pozitivní/částečná kompromitace databáze
- Částečná kompromitace
  - Informace o entitě i když neznáme konkrétní hodnotu

Protiopatření ve statistických databázích I.

- Omezení dotazu
  - Např. i sledování předchozích dotazů
- Úmyslná změna zdrojových dat
  - Např. orig. hodnoty nahrazeny novým vzorkem se stejným rozložením pravděpodobnosti hodnot
- Úmyslná změna výsledku dotazu
  - Např. zaokrouhlování,
- Cílem je zabránit situacím, kdy je možné získat informace o jedné entitě.

Protiopatření ve statistických databázích II. – *Náhodný výběr*

Každý dotaz je zodpovězen na základě vyhodnocení náhodně vybraných záznamů ze všech existujících záznamů.

- Kontrola překrytí množiny záznamů u vícenásobných dotazů na tutéž informaci.
  - Má zabránit situaci, kdy několik uživatelů databáze začne spolupracovat.
- Technika nyní používaná v americké databázi údajů ze sčítání lidu.

Protiopatření ve statistických databázích III. – *Minimální rozsah dotazu*

- Minimum celkového počtu záznamů použitých pro tvorbu odpovědí.

nebo

- Minimum počtu záznamů použitých pro tvorbu odpovědí na každou část dotazu.

Protiopatření ve statistických databázích III. – *Perturbační (zmatečné) techniky*

Přidání pseudonáhodného „šumu“:

- Odpovědi konzistentní, ale získání spolehlivé odpovědi na sérii podobných dotazů není možné.

1. K záznamům zahrnutým pro vyhodnocení dotazů se přidají další náhodně vybrané podobné záznamy
2. Vypočtená hodnota nebo mezihodnoty jsou zaokrouhlovány nebo mírně pozměněny.

- Podle některých definic zahrnují *náhodný výběr*.

De-anonymizace uživatelů

- Narayanan a Shmatikov (2008)
  - Huge de-anonymization of large sparse datasets (ACM)
- Databáze hodnocení filmů
  - Databáze zpřístupněna „anonymizovaně“
- Uživatel hodnotí filmy (filmů jsou stovky) na škále 1-10
- Uživatele se podařilo de-anonymizovat – spojit se skutečnou identitou pokud:
  - Víme jeho hodnocení pro 5-8 filmů

Informační bezpečnost ve zdravotnictví

PV080

Vašek Matyáš

Zdravotnictví a bezpečnost

SECURITY

- Vlastnost prvku (např. IS), který je na určité úrovni chráněn proti ztrátám nebo také stav ochrany (na určité úrovni) proti ztrátám.

SAFETY

- Předpoklad, že při specifikovaných podmínkách nedojde ke stavu ohrožení lidského života, zdraví, hodnot a prostředí.

Ohrožení života „počítačem“ v medicíně?

PŘÍMO

- Tyto případy jsou velmi výjimečné, spíše extrémní.
- Např. chyba v programu způsobí zvýšení dávek ozáření, kterému pak pacient podlehl.

NEPŘÍMO

- Častější případy.
- Počítač nebo jím řízený přístroj dodají chybné výsledky analýzy, na jejichž základě lékař stanoví chybný léčebný postup.

Důvěryhodnost a důvěrnost

- Podvržená data – autentizace.
- Rukopis laboranta, razítko ap.

- Digitální podpis – integrita, autenticita dat.
- Prokazatelná zodpovědnost.

- Mlčenlivost - osobní zdravotní informace získané při lékařském výkonu.

- *Lékař, pacient, sestra, vedoucí ústavu, zdrav. pojišťovna.*

Bezpečnost v klinických informačních systémech

- Víceméně roztříštěné úsilí při tvorbě směrnic a pravidel. (Bezp. politika!)

- "*Security in Clinical Information Systems*", British Medical Association (BMA), 1996

BMA-1

- Doktor může otevřít nový záznam, kde je uveden jen on a pacient na seznamu řízení přístupu.

- Pokud je pacient jen na speciálním vyšetření, pak může doktor na seznam zařadit i jeho ošetřujícího lékaře.

BMA-2

- Právě jeden z lékařů na seznamu řízení přístupu musí být označen jako odpovědný a pouze on může seznam měnit a může k němu přidávat jen odborné zdravotnické pracovníky.

BMA-3

- Odpovědný lékař musí pacientovi sdělit, kdo je na seznamu řízení přístupu při vytvoření nového záznamu, při jakýchkoliv změnách a kdykoliv je odpovědnost za záznam předávána jinému lékaři.
- Pacientův souhlas musí být výslovný, s výjimkou řešení nouzových stavů a specifikovaných statutárních případů.

BMA-4

- Nikdo nesmí mít možnost smazat klinické informace, dokud neuplynula předepsaná doba pro jejich úschovu.

BMA-5

- Všechny přístupy ke klinickým záznamům musí být zaznamenány s udáním informací kdo a kdy se záznamem pracoval. Auditní záznam všech mazání musí být neustále udržován.

BMA-6

- Informace ze záznamu A mohou být připojeny k záznamu B tehdy a jen tehdy, když seznam řízení přístupu záznamu B je obsazen v seznamu pro A.

BMA-7

- Musí být zavedena účinná opatření proti agregaci osobních zdravotních informací.
- Pacienti, k jejichž seznamu řízení přístupu má být přidána další osoba, musí být zvlášť upozorněni, pokud již tato osoba má přístup ke zdravotním informacím velkého množství lidí.

BMA-8

- Počítačové systémy, které pracují s osobními zdravotními daty, musí mít subsystém, který efektivně prosazuje výše uvedené principy. Účinnost tohoto subsystému musí být podrobena hodnocení nezávislými experty.

Polosemestrální písemka!

9. listopadu

Více informací 26. října